A stationary unbiased finite sample ARCH-LM test procedure

Sjölander, Pär

Submitted Manuscript

A Stationary Unbiased Finite Sample ARCH-LM Test Procedure

Journal:	*Applied Economics*
Manuscript ID:	APE-07-0240.R1
Journal Selection:	Applied Economics
Date Submitted by the Author:	13-Oct-2008
Complete List of Authors:	Sjölander, Pär
JEL Code:	C00 - General < C0 - General < C - Mathematical and Quantitative Methods, C10 - General < C1 - Econometric and Statistical Methods: General < C - Mathematical and Quantitative Methods, C15 - Statistical Simulation Methods|Monte Carlo Methods < C1 - Econometric and Statistical Methods: General < C - Mathematical and Quantitative Methods, C32 - Time-Series Models < C3 - Econometric Methods: Multiple/Simultaneous Equation Models < C - Mathematical and Quantitative Methods
Keywords:	ARCH-LM, GARCH, Non-negativity constraints, Stationarity constraints

Editorial Office, Dept of Economics, Warwick University, Coventry CV4 7AL, UK

A Stationary Unbiased Finite Sample ARCH-LM Test Procedure

Pär Sjölander*, Ph.D.
Jönköping University
P.O. Box 1026, SE-551 11, Jönköping
Sweden

Abstract

Engle's (1982) ARCH-LM test is the standard test to detect autoregressive conditional heteroscedasticity. In this paper, Monte Carlo simulations are used to demonstrate that the test's statistical size is biased in finite samples. Two complementing remedies to the related problems are proposed. One simple solution is to simulate new unbiased critical values for the ARCH-LM test. A second solution is based on the observation that for econometrics practitioners, detection of ARCH is generally followed by remedial modeling of this time-varying heteroscedasticity by the most general and robust model in the ARCH family; the GARCH(1,1) model. If the GARCH model's stationarity constraints are violated, as in fact is very often the case, obviously, we can conclude that ARCH-LM's detection of conditional heteroscedasticity has no or limited practical value. Therefore, formulated as a function of whether the GARCH model's stationarity constraints are satisfied or not, an unbiased and more relevant two-step ARCH-LM test is specified. If the primary objectives of the study are to detect and remedy the problems of conditional heteroscedasticity, or to interpret GARCH parameters, the use of this paper's new two-step procedure, 2S-UARCH-LM, is strongly recommended.

JEL classification: G00, C10, C15, C32.
Keywords: ARCH-LM, GARCH, Non-negativity Constraints, Stationarity Constraints.

* Correspondence to Pär Sjölander. E-mail address: Par.Sjolander@ihh.hj.se.

I would gratefully like to acknowledge the financial support provided by Sparbankernas Forskningsstiftelse.

1. Introduction

Engle's (1982) ARCH-LM test is the standard approach to detect autoregressive conditional heteroscedasticity.[1] However, according to previous research by for instance Engle, Hendry, and Trumble (1985), Hong and Shehadeh (1999), Luukkonen, Saikkonen, and Teräsvirta (1988), Diebold and Pauly (1989), and Bollerslev and Wooldrige (1992), the actual statistical size of the ARCH-LM test is generally less than its nominal size in finite samples. This implies that the nominal size of the test tends to overestimate the true probability of a type-1 error in finite samples. Consequently, if this is the case, there should be a good potential to adjust the statistical size of the original ARCH-LM test to increase the statistical power of this test. However, in this paper Monte Carlo simulations reveal that in practice the actual statistical sizes for medium and small sample sizes are higher than what is stipulated by the nominal statistical sizes, given that we rule out the irrelevant cases which exhibit violated non-stationarity constraints[2] in the GARCH model. Thus, when we apply a more relevant and realistic Monte Carlo simulation approach that replicates the factual test procedure that practitioners in fact follow (which includes ruling out cases where the stationarity constraints are violated), we find severe size overrejection for medium and low sample sizes.[3] For these reasons, new critical values are simulated to remedy the problems in both of the above mentioned approaches.[4]

As a consequence, a new test named the Two-Step Unbiased ARCH-LM (2S-UARCH-LM) test is constructed in this paper. In contrast to all previous ARCH tests, this new test takes into account whether or not we can remedy the autoregressive conditional heteroscedasticity problem that is detected by the applied ARCH test. If the primary objectives of the study are to detect and remedy the problems of conditional heteroscedasticity (e.g. to obtain more efficient standard errors by modeling a GARCH model), or to interpret GARCH parameters (and make a statement about financial risks), it is of crucial importance that the stationarity constraints of the GARCH estimates are not violated (for instance due to negative variances). The new test procedure only declares that relevant autoregressive conditional heteroscedasticity effects are identified when the ARCH test is significant <u>and</u> the stationarity constraints are satisfied in a GARCH(1,1) framework. If not both of these conditions are satisfied this new test does not define this series as a relevant autoregressive conditional heteroscedasticity process. Only if both conditions are simultaneously satisfied we can know that there is a remedy to the problem, and solving the noise problem in the residuals is usually the sole purpose behind the entire exercise of ARCH-detection tests.

If the estimated GARCH coefficients exhibit negative or explosive variances, this GARCH model is not useful despite that the ARCH-LM test is significant. Thus, this violation of the stationarity constraints is also a warning that a GARCH(1,1) model is not suitable for solving the noise problems in this model's residuals (often due to misspecification problems). In summary, the major reasons for why 2S-UARCH-LM should be applied is because it

---

[1] In this paper the name ARCH-LM is used, regardless of whether the test's χ^2- or F-versions are referred to. ARCH-LM is, despite its small-sample biasedness, the standard test to detect autoregressive conditional heteroscedasticity in all sample sizes (see for instance, Hodge (2005), Karadas and Öğünç (2005), and Weymark (1999)).

[2] These so-called "stationarity constraints" are presented in Table 1. Notice that these restrictions include both constraints regarding non-stationarity of the GARCH coefficients as well as non-negativity constraints for these coefficients.

[3] See the statistical size of the 2S-ARCH-LM test in Table 4.

[4] See the unbiased critical values of the UARCH-LM test in Table 5 (that adjusts the size of the ARCH-LM test), and the unbiased critical values of the 2S-UARCH-LM test in Table 5 (that adjusts the size of the 2S-ARCH-LM test).

exhibits good size and power properties and since it takes into consideration whether the GARCH constraints are satisfied or not.

In previous research we can actually find some unbiased alternatives[5] to Engle's ARCH-LM test that have been proposed by for instance Bera and Higgins (1992), Gregory (1989), Lee (1991), Lee and King (1993), Robinson (1991), and Hong and Shehadeh's (1999). However, the main reason why neither of the previous tests constitute any relevant substitute to the new test in this paper is because none of them takes into consideration whether the GARCH model's stationarity constraints are satisfied or not. For instance, these traditional approaches do not distinguish whether the variances are negative or whether there are very explosive coefficients in a series. Therefore, these tests are not directly comparable or any valid substitute to the new 2S-UARCH-LM test in this paper. Moreover, the impact factors of these traditional ARCH tests are very low compared to the ARCH-LM test which still is the undisputed number one standard test to detect autoregressive conditional heteroscedasticity. Accordingly, these complicated alternative methods are generally not applied by the average practitioner in economics which implies that these contributions have limited utility in practice. These methods involves for instance fairly complicated kernel-based weighting schemes based on frequency-domain approaches which is executed by programming in Gauss, MatLab, or some other flexible environment. One may also argue that there are some other problems and limitations in these tests, however, the main reason why the ARCH-LM test still is the standard test (despite its size problems) appears to be due to its simplicity and since it is included in many user-friendly program packages.

Consequently, this paper aims to introduce a simple and user-friendly unbiased procedure (based on new critical values) to detect and remedy conditional heteroscedasticity problems. This new approach can be applied by practitioners in economics with fairly basic skills in statistics and with no knowledge in programming. However, the unique contribution of this paper and the most important reason to apply this new ARCH-detection method (2S-UARCH-LM) is that it exhibit good size and power properties and that it, unlike previous methods, takes into account whether the GARCH stationarity constraints are satisfied or not.

2. The specifications of the ARCH tests

From the viewpoint of inference, neglecting ARCH effects may lead to arbitrary large losses in asymptotic efficiency (Engle, 1982) and cause overrejection of standard tests for autocorrelation in the conditional mean (see Taylor, 1984 and Diebold, 1987). Engle's (1982) ARCH-LM test statistic is still the most commonly applied standard test to detect autoregressive conditional heteroscedasticity. It is computed from an auxiliary test regression, and the null hypothesis is that there is no existing ARCH up to order q in the residuals (e_t). It is asymptotically locally most powerful if the true alternative is ARCH(q). The null hypothesis of no ARCH(q) is examined by running the following regression.

$$(1) \qquad e_t^2 = \hat{\delta}_0 + \sum_{s=1}^{q} \hat{\delta}_s e_{t-s}^2 + \upsilon_t$$

[5] Most of these mentioned alternative tests are not directly comparable to Engle's ARCH-LM test. These tests usually only test some alternative properties that are very related to the ARCH-LM test, but they definitely not direct substibutes to the ARCH-LM test.

Thus, the squared residuals are regressed on a constant and lagged squared residuals up to order q. There are two commonly applied versions of the test. One is the ARCH-LM test statistic that is computed by the number of observations multiplied by the R^2 from the regression in Equation 1. A lagrange multiplier interpretation can be given the test statistic and it is asymptotically distributed as a $\chi^2(p)$ random variable. However, the F-statistic version of the test is an omitted variable test for the joint significance of all lagged squared residuals. From previous research we know that the exact finite sample distribution of the F-statistic under the null hypothesis is not known.

Usually, detection of autoregressive conditional heteroscedasticity is followed by the modeling of this time-varying heteroscedasticity. In the family of ARCH, the most robust and most commonly used modeling procedure is the GARCH(1,1) model (Bollerslev, 1986). It is well known that the stationarity constraints very often are not satisfied for the estimated GARCH(1,1) models (see Table 1). Satisfied stationarity constraints in GARCH(1,1) are crucial in order to avoid meaningless processes, such as models with negative variances or very explosive processes. If the ARCH-LM test concludes that there is a conditional heteroscedasticity problem but the GARCH(1,1) model estimates coefficient values which implies negative variances, then there is no simple remedy to the problem.

Prior to the presentation of the new test procedure, it is necessary to present some fundamental concepts on the topic of GARCH models. In Equation 2 a GARCH(1,1) process is presented.

$$(2) \qquad \varepsilon_t = \eta_t \sqrt{h_t}\,, \; h_t = \omega + \alpha\, \varepsilon_{t-1}^2 + \beta\, h_{t-1}$$

Obviously this data generating process (DGP) is also the test specification of an empirical GARCH(1,1) model, where the GARCH coefficients are interpreted as follows. $\hat{\omega}$ is the estimated weighted average of the long-term average. $\hat{\alpha}$ is the ARCH component that estimates the relation to the conditional volatility observed in the previous period, while $\hat{\beta}$ (the GARCH component) is the estimated forecasted variance from the last time period. In the GARCH(1,1) process, it is roughly accurate to say that α determines the degree of volatility of the variance process, while $\sum(\alpha+\beta)$ determines its persistence.[6]

If the residuals from an arbitrary OLS regression model exhibit the GARCH problems observed in Equation 2, it is a well known fact that this heteroscedasticity can cause problems in the analysis. The regression coefficients are still unbiased, but the confidence intervals will be too tight, giving a false sense of precision. For large sample sizes, so-called Robust Standard Errors (or quasi-maximum likelihood, QML, covariances and standard errors) by Bollerslev and Wooldridge (1992) can improve the estimates of standard errors in the presence of conditional heteroscedasticity. This option is advised when the residuals in the DGP are not conditionally normally distributed. In situations when the assumption of conditional normality do not hold, the (G)ARCH parameter estimates will still be consistent

[6] According to for instance Lamont, Lumsdaine and Jones (1996) and Gallo and Pacini (1997) $\sum(\alpha+\beta)$ determines the GARCH model's persistence.

Editorial Office, Dept of Economics, Warwick University, Coventry CV4 7AL, UK

given that the model is correctly specified. For reasons related to risk aversion it is advised that QML is applied for GARCH(1,1) models as well.[7]

In order to be able to make the above interpretations regarding the GARCH coefficients, it is necessary to examine whether the stationarity constraints in Table 1 are satisfied or not. Since h_t in Equation 2 represents the conditional variance, it must always be strictly positive. It is well known that the estimated volatility coefficients often exhibit negative or explosive estimates. Obviously, negative variances cannot be interpreted from an intuitive perspective and therefore such a model is completely misleading and useless. The constraints for the GARCH(1,1) parameters are as follows; the constant $\omega>0$, the ARCH parameter $0\leq\alpha<1$, GARCH parameter $0\leq\beta<1$, $\sum(\alpha+\beta)<1$, and the unconditional variance UV>0 (see e.g. Franses, Dijk and Lucas (2004), Gourieroux (1997), Kim and Schmidt (1993), or Li, Ling, and McAleer (2003)). Thus, if

$$UV = \left[\frac{\omega}{1-\alpha-\beta}\right] > 0 \qquad (3)$$

is satisfied, then the UV is well defined, which is implied by the constraints $\omega>0$ and $\sum(\alpha+\beta)<1$. The constraint $\omega=1-\alpha-\beta$ implies that the UV is equal to 1 for all simulated processes. These constraints are summarized below in Table 1.

Table 1 – Stationarity constraints of the DGPs

Factor	Symbol	Constraint
ARCH parameter constraint	α	$0 \leq \alpha < 1$
GARCH parameter constraint	β	$0 \leq \beta < 1$
Conditional variance constraint	h	$h > 0 \quad \forall t$
Persistence constraint	δ	$\sum(\alpha+\beta) < 1$
Unconditional variance constraint	σ^2	If $\sigma^2 = (\omega / (1 - \alpha - \beta)) > 0$ is satisfied, then the σ^2 is well defined, which is implied by the constraints $\omega > 0$ and $\sum(\alpha+\beta) < 1$.

Notice that the so-called stationarity constraints in Table 1 includes both constraints regarding non-stationarity of the GARCH parameters as well as non-negativity constraints for these parameters. Therefore, in the literature it is standard to refer to these GARCH constraints in Table 1 as "stationarity constraints" despite that this concept also includes constraints regarding parameter non-negativity.

One of the main contributions of this paper is to present a completely new, but in practice more relevant, form of definitions for statistical size and statistical power analysis. Rejection of the null hypothesis is only registered when the stationarity constraints are satisfied so that it is possible to remedy the problem. Counting how many times ARCH-LM detects conditional heteroscedasticity simultaneously as the estimated parameters are negative or explosive is not relevant in reality. For instance, in financial economics or for option pricing models, detection of conditional heteroscedasticity is not meaningful unless the stationarity

[7] Unless the QML approach is applied, the covariance matrix estimates will not be consistent which would result in incorrect standard errors. The QML correction does not affect the estimated coefficients only the estimated standard errors. Consequently, this adjustment does not affect the executed simulation studies since new simulated critical values are applied. However, when estimating GARCH(1,1) coefficients this correction adjusts the standard errors. See for instance Chan and McAleer (2003).

constraints are satisfied, since prices cannot be derived from a model with negative variances or explosive processes. Therefore a new form of relevant size and power concept is formulated in this paper.

It is possible to constrain the estimated coefficients not to go beyond certain values. However, it is well known that such restrictions often lead to misleading values of the estimates. For instance, the EGARCH model proposed by Nelson (1991) specifies the left-hand side of the conditional heteroscedasticity equation as the log of the conditional variance. Therefore, this implies that the leverage effect is exponential, rather than quadratic, and that forecasts of the conditional variance are guaranteed to be nonnegative. However, obviously, since the true DGP in the simulation experiment is generated from the most general and robust process, GARCH(1,1), estimating these processes with something other than a GARCH(1,1) model (e.g. an EGARCH or an IGARCH model) would be totally incorrect and irrational.

3. The ARCH-detection procedures

The following test procedures are defined or created in this paper and later applied in the Monte Carlo (MC) simulation section of this study:[8]

- **ARCH-LM** (Used as a benchmark test in the MC study)[9]: Engle's (1982) test can only be used to detect whether there exist ARCH effects or not. The test is biased in finite samples, and does not take into account whether the stationarity constraints are satisfied or not.

- **2S-ARCH-LM** (Used as a benchmark test in the MC study)[10]: This traditional two-step approach can detect whether there exists ARCH noise in the residuals that is possible to remedy by using a stationary GARCH(1,1) model. This is an replication of the actual statistical power that a researcher has to face if the objective is to detect and solve the problem of conditional heteroscedasticity in the residuals. This two-step approach replicates the current standard procedure that is followed by most econometric practitioners. Therefore, this test procedure can be used as a benchmark in comparison with the new suggested 2S-UARCH-LM method in this paper. In this paper it is illustrated that 2S-ARCH-LM is biased in finite samples, but it rules out models where the stationarity constraints are violated.

- **UARCH-LM** (New test): This new test can only detect whether there exist ARCH effects or not. In contrast to the 2S-UARCH-LM test below, UARCH-LM does not take into account whether the stationarity constraints are satisfied or not. In this paper it is illustrated that the lagrange multiplier (LM) test version of the test statistic (based on Equation 1) in fact is not χ^2-distributed in finite samples.[11] Moreover, the

[8] 2S-(U)ARCH-LM test stands for Two-stage (unbiased) autoregressive conditional heteroscedasticity lagrange multiplier test, since in the first step the ARCH-LM is executed, and in the next step we evaluate whether the stationarity constraints are satisfied. Since the 2S-ARCH-LM procedure exhibits severe size-problems, new remedial critical values are simulated for 2S-ARCH-LM (2S-ARCH-LM is simply a replication of the currently most common procedure that is applied by practitioners, and is for practical reasons given a name in this paper since it is used as a benchmark model that is compared with the new approach). Besides new critical values, exactly the same procedure is used in the new test. However, since the new critical values make the 2S-ARCH-LM unbiased, the test new procedure is denoted 2S-UnbiasedARCH-LM or 2S-UARCH-LM.

[9] In this paper this ARCH-LM test is applied as a benchmark test in order to be able to compare the performance of the new UARCH-LM test.

[10] In this paper this the 2S-ARCH-LM test is applied as a benchmark test in order to be able to compare the performance of the new 2S-UARCH-LM test.

[11] This is demonstrated in the Monte Carlo simulations in Table 4.

omitted variable test statistic for the joint significance of all lagged squared residuals (the F-test version of the ARCH test) is not F-distributed in finite samples.[12] Therefore new unbiased critical values for these test statistics are presented in Table 5.

- **2S-UARCH-LM** (New test)[13]: Unlike all previous tests, this new two-step test can detect whether there are significant ARCH effects in the residuals simultaneously as it can determine whether this noise is possible to remedy by using a stationary GARCH(1,1) model. Consequently, this approach requires that the stationarity constraints in Table 1 are satisfied. In this paper it is illustrated that the lagrange multiplier (LM) test version of the 2S-ARCH-LM test statistic (that is based on Equation 1) in fact is not χ^2-distributed in finite samples.[14] Moreover, the omitted variable test statistic for the joint significance of all lagged squared residuals (the F-test version of the ARCH test) is not F-distributed in finite samples.[15] Therefore new unbiased critical values for these test statistics are presented in Table 6. Unlike the critical values in Table 5 for the UARCH-LM test, these critical values takes into account whether the stationarity constraints are violated or not. This test is the main contribution of the paper since it is the only test that controls for whether the stationarity constraints in Table 1 are satisfied or not.

  In summary, the new 2S-UARCH-LM test is conducted as follows. First, based on Equation 1, we construct an ARCH test statistic. If this test do not reject the null hypothesis of no ARCH (based on the critical values in Table 6), we conclude that no autoregressive conditional heteroscedasticity is present in the tested series. However, if the ARCH test rejects the null hypothesis of no ARCH this is only a necessary, but not a sufficient condition to conclude that there are stationary and solvable autoregressive conditional heteroscedasticity effects. It is also necessary that the stationarity constraints from Table 1 are satisfied to conclude that we have detected stationary and relevant GARCH effects. If the GARCH coefficients in the data generating process violates the stationarity constraints in Table 1 there is no cure for the problem. If there is no remedy to the problem the entire exercise is pointless and this is also an indication that the true DGP probably does not follow a GARCH model, and consequently modeling this process using a GARCH model is probably a bad idea to solve the problem.

The following test hypotheses are examined for the studied tests in this paper:

Test hypotheses for ARCH-LM and UARCH-LM:

(i) Under the null hypothesis, ARCH up to order q cannot be detected.
(ii) Under the alternative hypothesis, ARCH up to order q can be detected.

---

[12] *Ibid.*
[13] Furthermore, observe that the above two-step approaches do not suffer from massignificance since there are if-statements that evaluate the constraints instead of p-values.
[14] This is demonstrated in the Monte Carlo simulations in Table 4.
[15] *Ibid.*

Test hypotheses for 2S-ARCH-LM and 2S-UARCH-LM:

(i) Under the null hypothesis, ARCH up to order q cannot be detected, OR ARCH up to order q can be detected simultaneously as the stationarity constraints in Table 1 are violated for the GARCH(1,1) model.
(If the stationarity constraints are not satisfied we have not detected a model that contains meaningful or realistic parameters, and there is no remedy to potential autoregressive conditional heteroscedasticity problems).

(ii) Under the alternative hypothesis, ARCH up to order q can be detected, AND the stationarity constraints in Table 1 are satisfied for the GARCH(1,1) model.
(If the stationarity constraints are satisfied we have detected a model that contains meaningful and realistic parameters, which can be applied to remedy the detected autoregressive conditional heteroscedasticity problems).

In summary, the main contribution of this paper is the new two-step approach that takes into account whether the stationarity constraints are satisfied or not. It is named the 2S-UARCH-LM test and the test procedure is summarized below.

Step 1: Estimate the ARCH detection test in Equation 1. Compare this 2S-UARCH-LM test statistic to the new unbiased simulated critical values found in Table 6.

(i): If the null hypothesis of no ARCH(q) is not rejected, conclude that there is no presence of autoregressive conditional heteroscedasticity in the tested series.

or

(ii): If the null hypothesis of no ARCH(q) is rejected, go to Step 2.

Step 2: Estimate a GARCH(1,1) model on the examined series in Step 1.

(i): If the stationarity constraints in Table 1 are not satisfied, conclude that there is no remedy to the ARCH problem. *(Furthermore, this violation is also an indication that the true DGP does not follow an autoregressive conditional heteroscedasticity process).*

or

(ii): If the stationarity constraints are satisfied, conclude the existence of a autoregressive conditional heteroscedasticity process, and that it is possible to remedy this noise problem by using a GARCH(1,1) model.

In this paper it is illustrated that it is important to take into consideration whether autoregressive conditional heteroscedasticity models contain stationary and meaningful coefficient estimates or not, when the rejection frequencies of a test are evaluated. The 2S-ARCH-LM and 2S-UARCH-LM approaches are much closer to the real-world situations that econometrics practitioners face, since it is not interesting to evaluate whether we have significant ARCH effects if these effects are based on e.g. negative variances or explosive GARCH parameters. Thus, this new 2S-UARCH-LM approach is suggested in order to obtain a more relevant size and power evaluation of the test's ability to detect and remedy autoregressive conditional heteroscedasticity problems.

4. The Monte Carlo simulation design

Monte Carlo simulations are applied in order to evaluate the statistical size and the statistical power properties of different forms of ARCH tests.

If the estimated statistical size is not equal to the nominal size this implies that a test is biased in the size. There are two categories of size-biased tests. Either the test rejects the true null hypothesis more often than what the nominal size stipulates (overrejection), or the test rejects fewer times than it should (underrejection). If the estimated size is unequal to the nominal size, it implies that we too often or too infrequently draw the conclusion that the series contains ARCH, given that there is no ARCH under the null hypothesis.

The Monte Carlo experiment has been executed by generating data from the DGP in Equation 2. Definitions of the variable names and the parameter sample space of Equation 2 are available in Table 2.

Table 2 – The varying factors used in the size and power simulations' DGPs

Factor	Symbol	Design
Nominal size	π_0	0.01, 0.05, 0.10.
Number of repetitions	N	50 000 for the sizes, 5 000 for the powers.
Number of observations	T	50, 100, 150, 200, 250, 500, 1 000, 1 500, 2 000, 2 500, 5 000.
ARCH parameter	α	0.00, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90 as long as the stationarity constraints are satisfied.
GARCH parameter	β	0.00, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90 as long as the stationarity constraints are satisfied.
The GARCH model's long-term average	ω	$\omega = 1 - \alpha - \beta$.
White noise process for the innovations	η	$\eta_t \sim iidN(0,1)$.

Furthermore, the nominal sizes of 1 %, 5 %, 10 % are chosen since they are the most commonly applied significance levels. 50 000 repetitions were selected for the size, and 5 000 replications were chosen for the power. The estimated error margin of the size ($\hat{\pi}$) is based on a 95 % binomially distributed confidence interval:

$$\hat{\pi} \pm Z\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{N}} \qquad (4)$$

This is an acceptable simulation error margin with a 95 % confidence interval for the size, and if the estimated size is outside this interval the test is defined as a size-biased.[16] The statistical size problem is important since a statistical size that is too low leads to reduced power for the test (that is, the probability of rejecting a false null hypothesis will be reduced if the actual size is lower than the stipulated nominal size). On the other hand, if the actual

[16] This is an evaluation approach applied in, for instance, Edgerton and Shukur (1999).

size is too high this leads to a highly misleading type-1 errors (that is, the probability of rejecting a true null hypothesis is too high). Therefore, in direct accordance with previous research in various areas of statistical literature, I consider an unbiased size to be a very important feature for a statistical test[17].

The initial starting value of h_t in the Monte Carlo simulations is set to $h_0=1$. The number of start-up values for the DGP (prior that the estimation measurement process is initialized) is 200, which should be considered as more than usual and therefore sufficient for this type of study.

5. Results and analysis

5.1 The stationarity constraints

As previously mentioned, a general problem with GARCH-errors in the data generating process is that even if we simulate true clean GARCH(1,1) processes, the stationarity constraints are often not satisfied for an estimated GARCH(1,1) model (see Table 3 below). In Table 3, based on simulated data generating processes, the "percentage share of satisfied stationarity constraints" measures how often the estimated GARCH-models satisfy all the stationarity constraints from Table 1.[18] Despite that the GARCH(1,1) model usually is considered to be the most general and robust model in the ARCH-family, it is remarkable how often the stationarity constraints are violated. For a sample size of 50 observations, the stationarity constraints are violated in at least the majority of the cases in Table 3. It is also illustrated that the GARCH(1,1) model very often fails to estimate pure ARCH-models or GARCH(p,0) models. In fact, when the true GARCH parameter (β) is zero the GARCH(1,1) estimates often exhibit negative betas (β).[19]

[17] The general decision rule in statistical hypothesis testing is that we usually only draw conclusions if a test is significant, while if it is insignificant we say that the test is inconclusive (which implies that we do not have enough statistical evidence to be able to reject the null hypothesis). Inconclusive does not imply that we believe in the null hypothesis, at least if we exclude unit root tests. Consequently, somewhat simplified, power problems may lead to too many inconclusive tests (with no decisions), while size problems leads to too many incorrect decisive conclusions. Since this is the usual practice in statistical hypothesis testing, it is necessary that the size is unbiased.

[18] If, in repeated simulations, the coefficient estimates of the GARCH(1,1) model always satisfy all the stationarity constraints in Table 1, then the "percentage share of satisfied stationarity constraints" is 100 (%). Therefore, optimally, this share should be 100 (%) since we only simulate pure *stationary* GARCH(1,1) data generating processes with no violated constraints, and then we should of course expect the estimated GARCH(1,1) models to obtain coefficient estimates that are stationary. However, if the estimation process does not work optimally this may result in a certain fraction of the estimated GARCH(1,1) models with coefficients that violates the stationarity constraints in Table 1.

[19] Moreover, there are many examples of high ARCH-component values (α) that can be found in for instance, Brooks *et al.* (2001), *Journal of International Financial Markets, Institutions and Money*, 11, p. 221, Kim K. and Schmidt P., (1993), "Unit Root Tests with Conditional Heteroscedasticity", *Journal of Econometrics*, 59, 287-300, Li W.K., Ling S., and McAleer M., (2003), "Estimation and Testing for a Unit Root Process with GARCH(1,1) Errors", *Econometrics Reviews*, 18, 722-729.

Table 3 – Percentage shares of satisfied stationarity constraints for different DGPs[20]

GARCH-parameters		T										
α	β	50	100	150	200	250	500	1 000	1 500	2 000	2 500	5 000
0.0	**0.0**	12.92	17.51	19.92	20.94	21.99	24.50	26.61	27.79	28.32	28.62	30.94
0.1	**0.0**	19.74	29.54	36.26	39.78	42.72	48.78	52.76	54.32	53.32	52.26	52.80
0.1	**0.1**	20.72	31.66	39.02	43.46	47.04	54.96	61.12	64.98	67.16	68.58	74.14
0.1	**0.2**	21.76	33.56	41.96	47.40	51.86	61.68	70.82	75.46	78.96	81.84	89.62
0.1	**0.3**	22.90	35.86	45.22	51.30	56.42	68.30	78.34	83.50	86.74	90.38	96.30
0.1	**0.4**	24.54	39.06	48.74	55.92	60.88	74.14	84.46	89.54	92.66	95.06	99.06
0.1	**0.5**	26.14	41.94	52.90	60.60	65.56	80.06	89.64	93.86	96.02	97.80	99.64
0.1	**0.6**	27.62	45.30	57.36	65.06	70.62	84.66	93.02	96.30	97.82	98.68	99.80
0.1	**0.7**	28.94	49.48	62.26	70.12	76.18	89.44	95.28	97.52	98.34	99.02	99.64
0.1	**0.8**	28.02	52.72	67.72	76.40	82.14	94.26	97.42	98.60	99.42	99.12	99.80
0.2	**0.0**	24.56	37.16	43.32	46.00	47.76	48.66	50.46	50.60	51.52	52.12	54.16
0.2	**0.1**	27.24	42.30	50.50	55.02	58.06	64.02	71.42	75.86	79.90	83.02	91.42
0.2	**0.2**	30.20	48.16	57.24	63.30	67.92	77.36	86.62	90.82	93.80	95.94	99.36
0.2	**0.3**	33.52	53.28	64.48	71.24	75.84	87.08	94.16	97.44	98.62	99.42	99.96
0.2	**0.4**	36.48	59.12	71.48	77.94	82.68	94.06	98.28	99.36	99.84	99.86	100.0
0.2	**0.5**	39.42	64.84	78.10	84.38	88.12	97.36	99.38	99.88	99.90	99.96	100.0
0.2	**0.6**	41.74	70.50	83.54	90.20	93.22	99.20	99.86	99.96	100.0	100.0	100.0
0.2	**0.7**	40.74	73.34	87.02	94.14	96.42	99.82	99.98	100.0	100.0	100.0	100.0
0.3	**0.0**	27.96	40.38	45.12	46.60	46.60	45.62	50.80	52.00	53.74	54.62	58.36
0.3	**0.1**	32.56	49.82	58.02	61.86	64.04	70.64	81.30	86.72	90.10	93.32	98.02
0.3	**0.2**	37.38	58.28	68.88	74.46	77.74	87.44	94.98	98.12	98.96	99.66	100.0
0.3	**0.3**	41.88	66.46	77.58	83.24	86.84	96.00	99.20	99.82	99.96	99.98	100.0
0.3	**0.4**	45.60	73.70	85.18	90.48	93.20	98.72	99.92	99.98	100.0	100.0	100.0
0.3	**0.5**	47.84	78.74	90.38	94.90	97.04	99.82	99.98	100.0	100.0	100.0	100.0
0.3	**0.6**	47.34	78.58	90.18	94.92	97.28	99.68	100.0	100.0	100.0	100.0	100.0
0.4	**0.0**	30.06	41.60	44.62	44.98	44.32	43.64	52.32	54.28	56.22	57.08	62.44
0.4	**0.1**	36.22	54.20	63.08	67.12	70.36	77.86	88.56	93.38	95.54	97.42	99.76
0.4	**0.2**	42.04	66.04	76.20	81.94	85.46	94.32	98.68	99.70	99.92	99.98	100.0
0.4	**0.3**	46.54	73.92	85.22	90.58	93.32	98.68	99.90	99.98	100.0	100.0	100.0
0.4	**0.4**	49.70	79.42	90.58	95.00	97.36	99.82	99.98	100.0	100.0	100.0	100.0
0.4	**0.5**	48.74	77.00	86.56	91.36	93.70	98.32	100.0	100.0	100.0	100.0	100.0
0.5	**0.0**	30.12	40.72	42.52	43.04	42.98	42.40	54.52	56.58	58.14	59.84	66.34
0.5	**0.1**	38.84	57.90	66.58	71.98	75.48	84.62	93.98	97.38	98.66	99.48	100.0
0.5	**0.2**	44.72	70.08	80.30	87.42	90.26	97.38	99.66	99.94	100.0	100.0	100.0
0.5	**0.3**	47.42	75.62	86.32	92.58	95.24	99.56	99.98	100.0	100.0	100.0	100.0
0.5	**0.4**	47.82	73.36	81.66	86.34	89.64	95.88	100.0	100.0	100.0	100.0	100.0
0.6	**0.0**	29.80	38.84	40.92	41.34	41.86	41.68	55.98	58.94	60.84	62.56	69.90
0.6	**0.1**	38.64	59.10	68.76	75.66	79.50	90.10	97.06	99.00	99.62	99.98	100.0
0.6	**0.2**	43.56	68.98	79.68	86.80	90.64	98.42	99.90	99.98	100.0	100.0	100.0
0.6	**0.3**	45.20	67.94	76.62	81.22	85.24	93.18	100.0	100.0	100.0	100.0	100.0
0.7	**0.0**	27.68	36.08	38.36	39.24	40.56	40.56	58.92	62.12	64.46	66.46	75.48
0.7	**0.1**	37.48	56.86	66.78	74.72	79.72	92.28	98.82	99.74	99.94	100.0	100.0
0.7	**0.2**	41.32	61.08	69.42	75.60	79.88	89.54	99.96	100.0	99.98	99.98	100.0
0.8	**0.0**	25.24	32.16	33.50	34.90	37.26	38.44	63.70	68.00	70.94	73.42	82.50
0.8	**0.1**	35.52	50.12	57.64	65.32	69.64	83.56	99.58	99.90	100.0	100.0	100.0

The first column represents the ARCH-parameters, and the second stands for the GARCH-parameters. The table exhibits results for 50, 100, 150, 200, 250, 500, 1 000, 1 500, 2 000, 2 500, and 5 000 observations. 5 000 repetitions are executed for the power, and 50 000 replications for the size.

[20] The stationarity constraints are presented in Table 1.

We can also see that for low values of the ARCH-parameter (α), the stationarity constraints are often violated. Furthermore, as expected, the higher the number of observations the more likely is the GARCH(1,1) model to satisfy the stationarity constraints. However, in some cases for low (G)ARCH magnitudes not even extremely high samples sizes, for instance 5 000 observations, are not enough. For example, we can see that an estimated GARCH(1,1) model fails to satisfy the stationarity constraints, for a simulated GARCH process with α=0.1 and β=0.1, in as many as 25 % of the cases. Moreover, this study does not present simulations of near integrated processes or degenerate cases. However, for both of these types of processes the stationarity-constraint issue is even more problematic.

In summary, the most interesting finding from Table 3 is that the stationarity constraints are very often not satisfied. If the stationarity constraints fails to be satisfied, there is no remedy to potential autoregressive conditional heteroscedasticity problems even if we would know that the true process in fact follows a GARCH(1,1) process.

5.2 The statistical sizes of the ARCH-LM, 2S-ARCH-LM, UARCH-LM, and the 2S-UARCH-LM tests

Below in Table 4, a Monte Carlo simulation size analysis for the four tests, based on 50 000 replications, is presented. The nominal sizes are set to 5 % and the estimated error margin of the size is based on the 95 % binomially distributed confidence interval presented in Equation 4. Based on this formula in Equation 4, if the actual size is too far apart from the nominal size, a test is defined as biased.

Observe that the 2S-ARCH-LM test is the benchmark procedure that is created to replicate the usual testing procedure that currently is the standard approach among most econometric practitioners, while the 2S-UARCH-LM test is the main new contribution in this paper.[21]

Table 4 – Statistical sizes of (U)ARCH-LM and 2S-(U)ARCH-LM (in percent)

	ARCH-LM		2S-ARCH-LM		UARCH-LM		2S-UARCH-LM	
T	F	χ^2	F	χ^2	F	χ^2	F	χ^2
50	2.93	3.03	5.99	6.15	5.09 (UB)	5.11 (UB)	5.18 (UB)	5.19 (UB)
100	3.75	3.81	6.69	6.80	4.85 (UB)	4.81 (UB)	4.97 (UB)	5.12 (UB)
150	3.98	4.01	6.84	6.87	4.91 (UB)	4.89 (UB)	5.06 (UB)	5.09 (UB)
200	4.17	4.20	6.55	6.59	4.83 (UB)	4.87 (UB)	5.17 (UB)	5.11 (UB)
250	4.29	4.31	6.63	6.66	4.97 (UB)	4.91 (UB)	4.89 (UB)	4.85 (UB)
500	4.46	4.47	6.34	6.35	5.06 (UB)	4.91 (UB)	4.83 (UB)	4.83 (UB)
1 000	4.70	4.71	5.88	5.89	5.05 (UB)	4.92 (UB)	4.85 (UB)	4.82 (UB)
1 500	4.64	4.65	5.49	5.50	5.07 (UB)	5.17 (UB)	5.12 (UB)	5.04 (UB)
2 000	4.85 (UB)	4.86 (UB)	5.45	5.46	5.11 (UB)	5.14 (UB)	4.89 (UB)	4.86 (UB)
2 500	4.85 (UB)	4.85 (UB)	5.22 (UB)	5.22 (UB)	4.92 (UB)	5.07 (UB)	5.13 (UB)	5.09 (UB)
5 000	4.96 (UB)	4.96 (UB)	5.00 (UB)	5.00 (UB)	4.94 (UB)	5.02 (UB)	5.08 (UB)	4.81 (UB)

(UB) stands for "UnBiased test", while the shaded areas symbolize biased tests. The above figures are presented in percentage terms for 5 % nominal sizes. Moreover, note that the error margins are different for every test.

[21] ARCH-LM is the (biased) benchmark test for situations when violations of the stationarity constraints are of no importance, while UARCH-LM is the same test but with new unbiased critical values.

Based on the 95 % binomially distributed confidence intervals computed from Equation 4, the ARCH-LM and the 2S-ARCH-LM test are always outside the error margins (i.e. biased tests) except for very large sample sizes, while the UARCH-LM and the 2S-UARCH-LM tests are always within the error margins (i.e. unbiased tests). Roughly speaking, the fewer the number of observations, the more severe size-problems are exhibited in Table 4.

The analysis of ARCH-LM's size demonstrates severe underrejection of the true null hypothesis with a size as low as 2.93 % for 50 observations. Thus, this rejection frequency is 42 percent *[=(100(2.93/5)-1)-100]* lower than what is stipulated by the nominal size. On the other hand, for 2S-ARCH-LM, the peak of the size problem seems to be located at 150 observations.[22] Despite that the true null hypothesis contains no conditional heteroscedasticity, the econometric practitioner finds conditional heteroscedasticity 36 % *[=(100(6.87/5)-1)-100]* more often than what is stipulated by the nominal size of 5 % (if the evaluation method controls for the violations of the stationarity constraints). Thus, this examination demonstrates that 2S-ARCH-LM in fact leads to a severe overrejection of the true null hypothesis. Nevertheless, one should note that there are many different factors that in an interactive manner affect the size, and therefore it is not obvious that the size should decrease or increase monotonously as a function of the number of observations.

The size problems for ARCH-LM and 2S-ARCH-LM are of such high magnitude (see Table 4) that it is necessary to create a remedy to the problem. The ARCH-LM test that is used to detect autoregressive conditional heteroscedasticity severely underrejects the true null hypothesis in finite samples, while an overrejection is present for the 2S-ARCH-LM test. Thus the result in Table 4 indicates that this standard approach (2S-ARCH-LM) in reality leads to a severe overrejection of the true null hypothesis in finite samples. The overrejection of the 2S-ARCH-LM test generally decreases as the number of observations increases to a large sample size. For the ARCH-LM test an analogous, but inverted, pattern is exhibited where a sequentially decreasing underrejection is observed as the number of observations increases.

The solution to the size-bias problem is not based on increasing the sample size to many thousands of observations, since these sample sizes are rarely available in for instance international economics, macro economics or regional economics. On the contrary, the solution is based on the UARCH-LM test and primarily on the 2S-UARCH-LM test which are evaluated in Table 4. According to the evaluation method, we can see that at least 2 000 observations are necessary in order to avoid bias for the ARCH-LM test while around 2 500 observations are necessary to obtain an unbiased 2S-ARCH-LM test. In direct contrast to these previous tests, the UARCH-LM test and the 2S-UARCH-LM test are always unbiased for all sample sizes.

[22] In some fields of economics 150 observations is a relevant, relatively large and not uncommon sample size, such as for instance in macro economics.

5.3 The new critical values of the UARCH-LM, and 2S-UARCH-LM tests

The following critical values in Table 5 and Table 6 should be used for the UARCH-LM and the 2S-UARCH-LM tests, at the 1 %, 5 %, or 10 % significance levels, in order to avoid bias.

Table 5 – Unbiased critical values of UARCH-LM

	UARCH-LM		UARCH-LM		UARCH-LM	
	1%		5%		10%	
T	**F**	**χ^2**	**F**	**χ^2**	**F**	**χ^2**
50	6.3322495	5.8178715	3.2217240	3.1433505	2.2387635	2.2279080
100	6.6047965	6.3112405	3.4896125	3.4378840	2.4091260	2.3992115
150	6.7266350	6.5198110	3.5281365	3.4923200	2.4671495	2.4594385
200	6.7233875	6.5675035	3.5900320	3.5615750	2.5141585	2.5076795
250	6.6248145	6.5040115	3.6286765	3.6050960	2.5634790	2.5576905
500	6.5472510	6.4881260	3.6600780	3.6479420	2.6006085	2.5974820
1 000	6.6426715	6.6119435	3.7564565	3.7498635	2.6497360	2.6480140
1 500	6.4996060	6.4801540	3.7331130	3.7288020	2.6246830	2.6235900
2 000	6.6342450	6.6189010	3.7864680	3.7830865	2.6580665	2.6571915
2 500	6.6456150	6.6332840	3.7841410	3.7814415	2.6617375	2.6610330
5 000	6.6169520	6.6108460	3.8319280	3.8305240	2.7010840	2.7007050

The table is based on 50 000 Monte Carlo simulations.

Table 6 – Unbiased critical values of 2S-UARCH-LM

	2S-UARCH-LM		2S-UARCH-LM		2S-UARCH-LM	
	1%		5%		10%	
T	**F**	**χ^2**	**F**	**χ^2**	**F**	**χ^2**
50	9.1804835	8.0071172	4.5053736	4.2862186	3.0715212	3.0057905
100	8.6168023	8.0769674	4.6293866	4.5096137	3.0743744	3.0413692
150	9.1731693	8.7518379	4.5827972	4.5047121	3.0813130	3.0591122
200	8.7003763	8.4169739	4.4912270	4.4356979	3.0427695	3.0269083
250	8.4040056	8.1932827	4.5066509	4.4617351	3.0322056	3.0196872
500	7.9623371	7.8683220	4.3710681	4.3503964	2.9488889	2.9432919
1 000	7.6366461	7.5937999	4.1738238	4.1647613	2.9130139	2.9103544
1 500	7.2803952	7.2548395	4.0300070	4.0245574	2.7302491	2.7289195
2 000	7.2549584	7.2359365	4.0284820	4.0243985	2.7089211	2.7079608
2 500	6.9601647	6.9463769	3.9165919	3.9135903	2.6650496	2.6643406
5 000	6.6473484	6.6411744	3.8589650	3.8575310	2.6115635	2.6112440

The table is based on 50 000 Monte Carlo simulations.

As expected, the critical values are equivalent for the F- and χ^2-versions test for 5 000 observations. The UARCH-LM test must be applied, instead of the ARCH-LM test, if the number of observations is lower than 2 000 observations. The 2S-UARCH-LM should be used instead of the 2S-ARCH-LM procedure, if the number of observations is lower than 2 500 observations. These recommendations are based on the Monte Carlo simulation results in Table 4.

Furthermore, in order to ascertain that the critical values in Table 5 and Table 6 in fact are correct, 50 000 simulations were conducted for the size with continuously randomly selected random seeds.

5.4 The power properties of the UARCH-LM, and 2S-UARCH-LM tests

Since the pure ARCH-LM test is biased, there is not much point in estimating the power of this test in this study.[23] Instead, the unbiased version of the test, UARCH-LM, is evaluated. UARCH-LM uses the new simulated critical values from Table 5, and the power of the test is presented in Table 7. This is the statistical power of detecting autoregressive conditional heteroscedasticity, regardless whether there is a possible solution to the problem or not.

Analogously, for the same reasons as above, the power of the 2S-UARCH-LM test is evaluated instead the biased 2S-ARCH-LM test. 2S-UARCH-LM uses the new simulated critical values from Table 6, and the power of the test is presented in Table 8. This is the statistical power of detecting autoregressive conditional heteroscedasticity, when we take into account whether there is a possible solution to the problem or not.

[23] If one would allow a test to be severely size-biased it is pointless to evaluate the power of this test. In fact it is very easy to construct a test with 100 % power if this test always rejects the null hypothesis. The cost would be that the power is misleading due to a deceptive type-1 error, and consequently this would not be a meaningful test.

Table 7 – Statistical power of the UARCH-LM test for different DGPs (in percent)

GARCH-parameters		T										
α	**β**	**50**	**100**	**150**	**200**	**250**	**500**	**1 000**	**1 500**	**2 000**	**2 500**	**5 000**
0.1	**0.0**	10.92	16.08	21.98	26.52	31.56	52.16	78.62	91.24	96.52	98.86	100.0
0.1	**0.1**	11.06	16.20	22.26	26.66	31.82	52.68	78.82	91.38	96.72	98.90	100.0
0.1	**0.2**	11.18	16.54	22.84	27.00	32.28	53.28	79.14	91.56	96.74	99.00	100.0
0.1	**0.3**	11.30	16.64	23.04	27.38	32.48	53.82	79.56	91.82	96.86	99.00	100.0
0.1	**0.4**	11.52	16.92	23.26	28.08	33.12	54.34	80.06	92.00	97.10	99.02	100.0
0.1	**0.5**	11.50	17.18	23.76	28.56	33.96	55.30	80.82	92.52	97.38	99.14	100.0
0.1	**0.6**	11.44	17.48	24.12	29.04	34.52	56.36	82.26	93.34	97.76	99.40	100.0
0.1	**0.7**	11.60	17.88	24.80	30.54	36.42	59.04	84.80	94.44	98.36	99.60	100.0
0.1	**0.8**	11.06	18.46	26.40	32.58	39.52	64.84	89.28	96.92	99.26	99.88	100.0
0.2	**0.0**	20.56	35.98	49.12	59.84	68.96	92.08	99.60	100.0	100.0	100.0	100.0
0.2	**0.1**	21.08	36.62	49.88	60.28	69.22	92.40	99.68	100.0	100.0	100.0	100.0
0.2	**0.2**	21.30	37.10	50.50	61.30	69.86	92.54	99.72	100.0	100.0	100.0	100.0
0.2	**0.3**	21.54	37.42	51.16	62.00	70.44	92.92	99.76	100.0	100.0	100.0	100.0
0.2	**0.4**	21.80	38.18	51.90	63.02	71.30	93.32	99.78	100.0	100.0	100.0	100.0
0.2	**0.5**	21.46	38.92	53.00	64.22	72.34	94.00	99.80	100.0	100.0	100.0	100.0
0.2	**0.6**	21.38	39.98	54.30	65.74	74.06	94.94	99.88	100.0	100.0	100.0	100.0
0.2	**0.7**	20.76	40.84	56.34	68.30	77.22	96.32	99.94	100.0	100.0	100.0	100.0
0.3	**0.0**	31.26	54.18	71.28	81.90	88.84	99.22	100.0	100.0	100.0	100.0	100.0
0.3	**0.1**	31.32	55.20	72.26	82.12	89.24	99.28	100.0	100.0	100.0	100.0	100.0
0.3	**0.2**	31.84	55.56	72.86	82.66	89.32	99.36	100.0	100.0	100.0	100.0	100.0
0.3	**0.3**	31.98	56.24	73.20	83.42	89.68	99.40	100.0	100.0	100.0	100.0	100.0
0.3	**0.4**	32.06	56.78	73.88	83.86	90.12	99.52	100.0	100.0	100.0	100.0	100.0
0.3	**0.5**	32.18	57.28	74.64	84.50	90.92	99.60	100.0	100.0	100.0	100.0	100.0
0.3	**0.6**	31.84	58.66	75.48	85.96	92.26	99.68	100.0	100.0	100.0	100.0	100.0
0.4	**0.0**	40.36	68.66	84.32	92.36	95.98	99.98	100.0	100.0	100.0	100.0	100.0
0.4	**0.1**	40.64	68.96	84.98	92.74	96.06	99.98	100.0	100.0	100.0	100.0	100.0
0.4	**0.2**	40.92	69.26	85.24	93.00	96.28	100.0	100.0	100.0	100.0	100.0	100.0
0.4	**0.3**	41.54	69.54	85.90	93.08	96.76	99.98	100.0	100.0	100.0	100.0	100.0
0.4	**0.4**	41.56	70.02	86.32	93.66	96.90	99.98	100.0	100.0	100.0	100.0	100.0
0.4	**0.5**	41.32	70.80	86.48	93.58	97.08	99.96	100.0	100.0	100.0	100.0	100.0
0.5	**0.0**	48.34	77.50	91.28	96.64	98.62	100.0	100.0	100.0	100.0	100.0	100.0
0.5	**0.1**	48.52	78.30	91.34	96.90	98.66	100.0	100.0	100.0	100.0	100.0	100.0
0.5	**0.2**	49.02	78.92	91.72	97.02	98.76	100.0	100.0	100.0	100.0	100.0	100.0
0.5	**0.3**	49.22	79.42	92.12	97.02	98.74	100.0	100.0	100.0	100.0	100.0	100.0
0.5	**0.4**	48.58	79.76	92.32	97.02	98.78	99.96	100.0	100.0	100.0	100.0	100.0
0.6	**0.0**	55.08	83.52	95.00	98.36	99.48	100.0	100.0	100.0	100.0	100.0	100.0
0.6	**0.1**	54.68	84.02	95.18	98.56	99.50	100.0	100.0	100.0	100.0	100.0	100.0
0.6	**0.2**	54.32	84.86	95.40	98.50	99.56	100.0	100.0	100.0	100.0	100.0	100.0
0.6	**0.3**	54.44	85.12	95.34	98.54	99.56	100.0	100.0	100.0	100.0	100.0	100.0
0.7	**0.0**	60.32	88.22	96.90	99.20	99.76	99.98	100.0	100.0	100.0	100.0	100.0
0.7	**0.1**	60.00	88.02	97.16	99.24	99.80	99.98	100.0	100.0	100.0	100.0	100.0
0.7	**0.2**	59.60	88.32	96.98	99.18	99.72	100.0	100.0	100.0	100.0	100.0	100.0
0.8	**0.0**	64.22	90.26	97.58	99.36	99.82	99.98	100.0	100.0	100.0	100.0	100.0
0.8	**0.1**	64.40	90.80	97.94	99.46	99.80	99.98	100.0	100.0	100.0	100.0	100.0

The first column represents the ARCH-parameters, and the second stands for the GARCH-parameters. The table exhibits results for 50, 100, 150, 200, 250, 500, 1 000, 1 500, 2 000, 2 500, and 5 000 observations. 5 000 repetitions are executed for the power estimations.

Table 8 – Statistical power of the 2S-UARCH-LM test for different DGPs (in percent)

GARCH-parameters		T										
α	β	50	100	150	200	250	500	1 000	1 500	2 000	2 500	5 000
0.1	**0.0**	12.36	16.72	22.17	27.15	30.71	48.71	74.93	89.05	95.59	98.48	100.0
0.1	**0.1**	12.45	17.31	22.55	28.07	30.91	50.00	76.04	90.12	96.40	98.73	100.0
0.1	**0.2**	12.32	17.52	23.36	28.14	31.89	51.01	77.01	91.00	96.45	98.87	100.0
0.1	**0.3**	13.10	18.01	23.44	28.11	33.07	51.89	77.51	91.41	96.72	98.96	100.0
0.1	**0.4**	13.04	17.77	23.68	28.36	33.08	52.77	78.55	91.50	96.78	98.98	100.0
0.1	**0.5**	13.24	18.17	23.44	28.91	33.68	52.99	79.37	92.09	97.22	99.15	100.0
0.1	**0.6**	13.40	18.28	24.20	29.66	33.81	54.50	81.12	93.09	97.59	99.31	100.0
0.1	**0.7**	13.55	18.88	24.90	30.69	35.21	56.98	83.69	94.31	98.33	99.58	100.0
0.1	**0.8**	12.85	18.97	26.43	32.49	38.67	62.49	88.37	96.88	99.22	99.88	100.0
0.2	**0.0**	20.77	34.39	47.00	57.91	64.57	89.48	99.50	99.96	100.0	100.0	100.0
0.2	**0.1**	21.07	35.65	47.33	59.11	65.62	90.16	99.56	99.97	100.0	100.0	100.0
0.2	**0.2**	21.72	36.92	48.67	59.34	66.64	90.98	99.68	100.0	100.0	100.0	100.0
0.2	**0.3**	21.96	37.61	49.78	60.39	68.22	91.69	99.73	100.0	100.0	100.0	100.0
0.2	**0.4**	21.98	37.55	49.97	60.94	68.60	92.11	99.74	100.0	100.0	100.0	100.0
0.2	**0.5**	22.73	37.60	50.22	61.96	69.36	92.44	99.76	100.0	100.0	100.0	100.0
0.2	**0.6**	23.00	38.16	51.16	63.37	71.55	93.51	99.86	100.0	100.0	100.0	100.0
0.2	**0.7**	20.72	38.07	52.98	65.01	73.84	95.61	99.94	100.0	100.0	100.0	100.0
0.3	**0.0**	30.54	49.83	66.22	77.77	85.45	98.51	100.0	100.0	100.0	100.0	100.0
0.3	**0.1**	31.45	51.02	67.87	78.98	86.91	98.70	100.0	100.0	100.0	100.0	100.0
0.3	**0.2**	31.14	51.89	68.79	80.31	87.68	98.97	100.0	100.0	100.0	100.0	100.0
0.3	**0.3**	31.14	53.05	69.81	80.51	87.70	99.13	100.0	100.0	100.0	100.0	100.0
0.3	**0.4**	31.75	53.62	70.27	81.01	87.96	99.25	100.0	100.0	100.0	100.0	100.0
0.3	**0.5**	31.40	53.67	70.88	81.85	88.56	99.40	100.0	100.0	100.0	100.0	100.0
0.3	**0.6**	30.12	53.12	71.06	82.66	89.72	99.58	100.0	100.0	100.0	100.0	100.0
0.4	**0.0**	37.13	62.07	79.78	89.37	93.86	99.91	100.0	100.0	100.0	100.0	100.0
0.4	**0.1**	38.32	64.39	81.26	90.58	95.17	99.92	100.0	100.0	100.0	100.0	100.0
0.4	**0.2**	38.77	65.63	82.47	91.34	95.34	99.98	100.0	100.0	100.0	100.0	100.0
0.4	**0.3**	39.02	65.83	82.54	91.43	95.56	99.98	100.0	100.0	100.0	100.0	100.0
0.4	**0.4**	38.55	64.92	82.23	91.47	95.71	99.98	100.0	100.0	100.0	100.0	100.0
0.4	**0.5**	36.93	63.97	82.09	91.29	95.37	99.92	100.0	100.0	100.0	100.0	100.0
0.5	**0.0**	43.43	71.91	87.72	94.84	97.30	100.0	100.0	100.0	100.0	100.0	100.0
0.5	**0.1**	44.34	73.51	88.56	95.36	98.15	100.0	100.0	100.0	100.0	100.0	100.0
0.5	**0.2**	44.54	74.20	89.32	95.74	98.29	100.0	100.0	100.0	100.0	100.0	100.0
0.5	**0.3**	43.78	73.87	89.37	96.05	98.28	100.0	100.0	100.0	100.0	100.0	100.0
0.5	**0.4**	43.66	72.93	88.64	95.32	98.10	99.96	100.0	100.0	100.0	100.0	100.0
0.6	**0.0**	47.72	78.17	91.64	97.24	98.90	100.0	100.0	100.0	100.0	100.0	100.0
0.6	**0.1**	49.22	79.56	92.29	97.91	99.14	99.98	100.0	100.0	100.0	100.0	100.0
0.6	**0.2**	49.13	80.40	92.90	97.95	99.14	100.0	100.0	100.0	100.0	100.0	100.0
0.6	**0.3**	48.36	79.19	92.38	97.49	99.13	99.98	100.0	100.0	100.0	100.0	100.0
0.7	**0.0**	51.66	82.71	94.42	98.47	99.61	100.0	100.0	100.0	100.0	100.0	100.0
0.7	**0.1**	52.77	84.10	95.00	98.66	99.62	99.98	100.0	100.0	100.0	100.0	100.0
0.7	**0.2**	51.45	84.09	94.64	98.65	99.55	100.0	100.0	100.0	100.0	100.0	100.0
0.8	**0.0**	55.86	85.82	95.70	98.74	99.73	100.0	100.0	100.0	100.0	100.0	100.0
0.8	**0.1**	55.12	85.91	96.11	99.17	99.74	99.98	100.0	100.0	100.0	100.0	100.0

The first column represents the ARCH-parameters, and the second stands for the GARCH-parameters. The table exhibits results for 50, 100, 150, 200, 250, 500, 1 000, 1 500, 2 000, 2 500, and 5 000 observations. 5 000 repetitions are executed for the power estimations.

As can be seen in Table 7, the power is low for lower number of observations and for lower magnitudes of ARCH-errors (α). To some limited extent we can observe that a higher persistence ($\alpha+\beta$) increases the power of the test, even if α definitely is the dominating factor in this context. For 1 000 or more observations, given that α is higher than 0.1, the power always reaches approximately 100 percent. In Table 8 it is evident that the power of the 2S-UARCH-LM test follows the same pattern as the UARCH-LM test. 2S-ARCH-LM exhibits higher power for the lowest spectrum of GARCH-parameters since it rules out unrealistic models that do not satisfy the stationarity constraints.

In summary, it is necessary to apply both of these tests, UARCH-LM and 2S-UARCH-LM, but under different circumstances. The UARCH-LM can replace the traditional and biased ARCH-LM test, since UARCH-LM is unbiased and more powerful. These tests can only *detect* conditional heteroscedasticity but cannot distinguish whether the stationarity constraints are satisfied or not. Consequently, it is possible that the detected autoregressive conditional heteroscedasticity is unfeasible to model or interpret due to negative or explosive coefficient estimates. We can only draw conclusions regarding whether there exist some noise problems in the residuals, but there may be no cure for the problem.

On the other hand, 2S-ARCH-LM is a two-step procedure created to mimic a realistic testing procedure that currently is the standard approach among most econometric practitioners. If the test procedure detects conditional heteroscedasticity for which there is no remedy, this finding has limited value for a practitioner. This is the reason why 2S-ARCH-LM is a more meaningful procedure in practice compared to the ARCH-LM or the UARCH-LM tests. However, the above simulations demonstrate that this 2S-ARCH-LM approach leads to biased size due to severe overrejection of the true null hypothesis. Therefore, the unbiased 2S-UARCH-LM test procedure is suggested in order to remedy this size-bias problem and due to the fact that it controls for violations of the GARCH coefficient stationarity constraints.

6. Conclusions

Monte Carlo simulations of sample sizes of 50 to 1 500 observations indicate that the statistical size is severely biased for ARCH-LM's tests. Consequently, the F- or the χ^2-distribution tables cannot be applied in finite samples for the respective versions of the test. We can see that it is necessary to use at least 2 000 observations to avoid bias for the ARCH-LM test. On the other hand, if an economics practitioner tests for autoregressive conditional heteroscedasticity effects using an ARCH-LM test with the restraint that the stationarity constraints should be satisfied within a GARCH(1,1) framework, 2 500 observations are necessary to expect an unbiased result.[24]

Since sample sizes of these very large dimensions usually are not available in many areas of economics, we cannot rely on the supply of large data sets to obtain unbiased test results. Therefore, two new test approaches are presented in this paper; the UARCH-LM test and the 2S-UARCH-LM test. If the sole purpose of the study is to detect the presence of ARCH in small or medium sized samples, the UARCH-LM test is proposed. However, if the aim is

[24] In this paper, this approach is illustrated by the benchmark procedure named 2S-ARCH-LM.

to detect and remedy the problem of autoregressive conditional heteroscedasticity, 2S-UARCH-LM is recommended.

Thus, the main contribution of this paper is the new test named the 2-Step Unbiased ARCH-LM (2S-UARCH-LM) test. In contrast to all previous ARCH tests, this new test takes into account whether or not we can remedy the autoregressive conditional heteroscedasticity problem that is identified by the applied ARCH-detection test. If the primary objectives of the study are to detect and remedy the problems of conditional heteroscedasticity (e.g. to obtain more efficient standard errors by modeling a GARCH model), or to interpret GARCH parameters (and make a statement about financial risks), it is of crucial importance that the stationarity constraints of the GARCH estimates are not violated (for instance due to negative variances). The 2S-UARCH-LM test procedure only declares that relevant autoregressive conditional heteroscedasticity effects are identified when the ARCH test is significant and the stationarity constraints are satisfied in a GARCH(1,1) framework. If not both of these conditions are satisfied this new test does not consider this series as an actual or relevant autoregressive conditional heteroscedasticity process. Only if both conditions are simultaneously satisfied we can recognize that there is a remedy to the problem, and solving the noise problem in the residuals is usually the sole purpose behind the entire exercise of ARCH-detection tests.

If we only take into account the traditional ARCH-LM test, it is obvious that the test severely underrejects the true null hypothesis. However, when we apply the 2S-ARCH-LM test which aims to replicate the realistic standard procedure which is utilized by most practitioners (including checking whether the stationarity constraints are satisfied) we find severe size overrejection for medium and low sample sizes. For example, under some circumstances, despite that the true null hypothesis contains no conditional heteroscedasticity, simulations demonstrate the econometric practitioner can be severely deluded since autoregressive conditional heteroscedasticity is identified 36 % *[=(100(6.87/5)-1)-100]* more often than what is stipulated by the nominal size of 5 %. Obviously, since GARCH models are used as a base in many investment decisions, and since the model often is applied to adjust misleading standard errors in many areas of economics, this is a serious and relevant problem.

In summary, if the purpose of the study is solely to detect conditional heteroscedasticity, this can be determined by using the UARCH-LM test. On the other hand, if the aim is to detect conditional heteroscedasticity with the requirement of satisfied stationarity constraints, the 2S-UARCH-LM test should be applied. Unlike any other tests, the 2S-UARCH-LM procedure controls for violations of the stationarity constraints, simultaneously as it exhibit considerably better size and power properties compared to Engle's standard test for conditional heteroscedasticity.[25]

[25] Engle's (1982) ARCH-LM test is the undisputed standard test for detection of autoregressive conditional heteroscedasticity.

References

Bera A K and Higgins M L, (1992), "A Test for Conditional Heteroskedasticity in Time Series Models", *Journal of Time Series Analysis*, 13, 501-519.

Bollerslev T, (1986), "Generalizes Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31, 307-327.

Bollerslev T and Wooldrige J M, (1992), "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models With Time-Varying Covariances", *Econometric Reviews*, 11, 143-172.

Brock W, Dechert D, and Scheinkman J, (1987), "A Test for Independence Based on the Correlation Dimension", unpublished manuscript, University of Wisconsin, Dept. of Economics.

Brooks R D, Faff R W, and Fry T R L, (2001), "GARCH modelling of individual stock data: the impact of censoring, firm size and trading volume", *Journal of International Financial Markets, Institutions and Money*, 11, 215-222.

Brooks C and Rew A G, (2002), "Testing for a Unit Root in a Process Exhibiting a Structural Break in the Presence of GARCH Errors", *Computational Economics*, 20, 157-176.

Chan F and McAleer M, (2003), "Estimating Smooth Transition Autoregressive Models with GARCH Errors in the Presence of Extreme Observations and Outliers", *Applied Financial Economics*, 13, 581-592.

Diebold F X and Pauly P, (1989), "Small Sample Properties of Asymptotically Equivalent Tests for Autoregressive Conditional Heterskedasticity", *Statistische Hefte*, 30, 105-131.

Edgerton D and Shukur G, (1999), "Testing Autocorrelation in a System Perspective", *Econometric Reviews*, 18(4), 343-386.

Engle R F, (1982), "Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of U.K. Inflation", *Econometrica*, 50, 987-1008.

Engle R F, Hendry D F, and Trumble D, (1985), "Small Sample Properties of ARCH Estimators and Tests", *Canadian Journal of Economics*, 18, 66-93.

Franses H F, Dijk D, and Lucas A, (2004), Short Patches of Outliers, ARCH and Volatility Modelling, *Applied Financial Economics*, 14, 221-231.

Gallo G M and Pacini B, (1998), "Early News is Good News: The Effects of Market Opening on Market Volatility", *Studies in Nonlinear Dynamics and Econometrics*, 2(4), 115-131.

Gourieroux C, (1997), "ARCH Models and Financial Applications", Springer-Verlag: New York.

Gregory A W, (1989), "A Nonparametric Test for Autoregressive Conditional Heterskedasticity: A Markov-Chain Approach", *Journal of Business and Economic Statistics*, 7, 107-115.

Hodge D, (2006), "Inflation and growth in South Africa", *Cambridge Journal of Economics,* 30(2), 163-180.

Hong Y and Shehadeh R D, (1999), "A New Test for ARCH Effects and Its Finite-Sample Performance", *Journal of Business and Economic Statistics*, 17(1), 91-108.

Karadaş E and Öğünç F, (2005), "An Analysis of Inflation Expectations of the Turkish Private Manufacturing Industry", *Statistics in Transition*, 7(2), 365-390.

Kim K and Schmidt P, (1993), "Unit Root Tests with Conditional Heteroscedasticity", *Journal of Econometrics*, 59, 287-300.

Lamont O, Lumsdaine R, and Jones C M, (1996), "Public Information and the Persistence of Bond Market Volatility", Working Paper 5446, National Bureau of Economic Research, Cambridge, MA.

Lee J H H, (1991), "A Lagrange Multiplier Test for GARCH Models", *Econometrics letters*, 37, 265-271.

Lee J H H and King M L, (1993), "A Locally Most Powerful Based Score Test for ARCH and GARCH Regression Disturbances", *Journal of Business and Economic Statistics*, 11(1), 17-27.

Li W K, Ling S, and McAleer M, (2003), "Efficient Estimation and Testing for Unit Root Processes with GARCH (1,1) Errors: A Monte Carlo Analysis", *Econometric Reviews*, 22, 179-202.

Luukkonen R, Saikkonen P, and Teräsvirta T, (1988), "Testing linearity in Univariate Time Series Models", *Scandinavian Journal of Statistics*, 15, 161-175.

Nelson D B, (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach", *Econometrica*, 59, 347-370.

Robinson P M, (1991), "Testing for Strong Serial Correlation and Dynamic Conditinoal heterskedasticty in Multiple Regression", *Journal of Econometrics*, 47, 67-84.

Weymark D N, (1999), "Heteroscedastic Exchange Rates and Time-Varying Optimal Intervention Rules", *The Canadian Journal of Economics / Revue canadienne d'Economique*, 32(5), 1174-1190.